The State of the PDS4 Information Model

Steve Hughes, Dan Crichton
Jet Propulsion Laboratory
California Institute of Technology

4th Planetary Data Workshop
June 18–20, 2019
Flagstaff, Arizona

Planetary Data Archives
Support Services, and Tools I
Wednesday, June,19 2019
8:45 a.m. HCCC Humphreys

Jet Propulsion Laboratory
California Institute of Technology

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Foundational Principles[1,2,3]

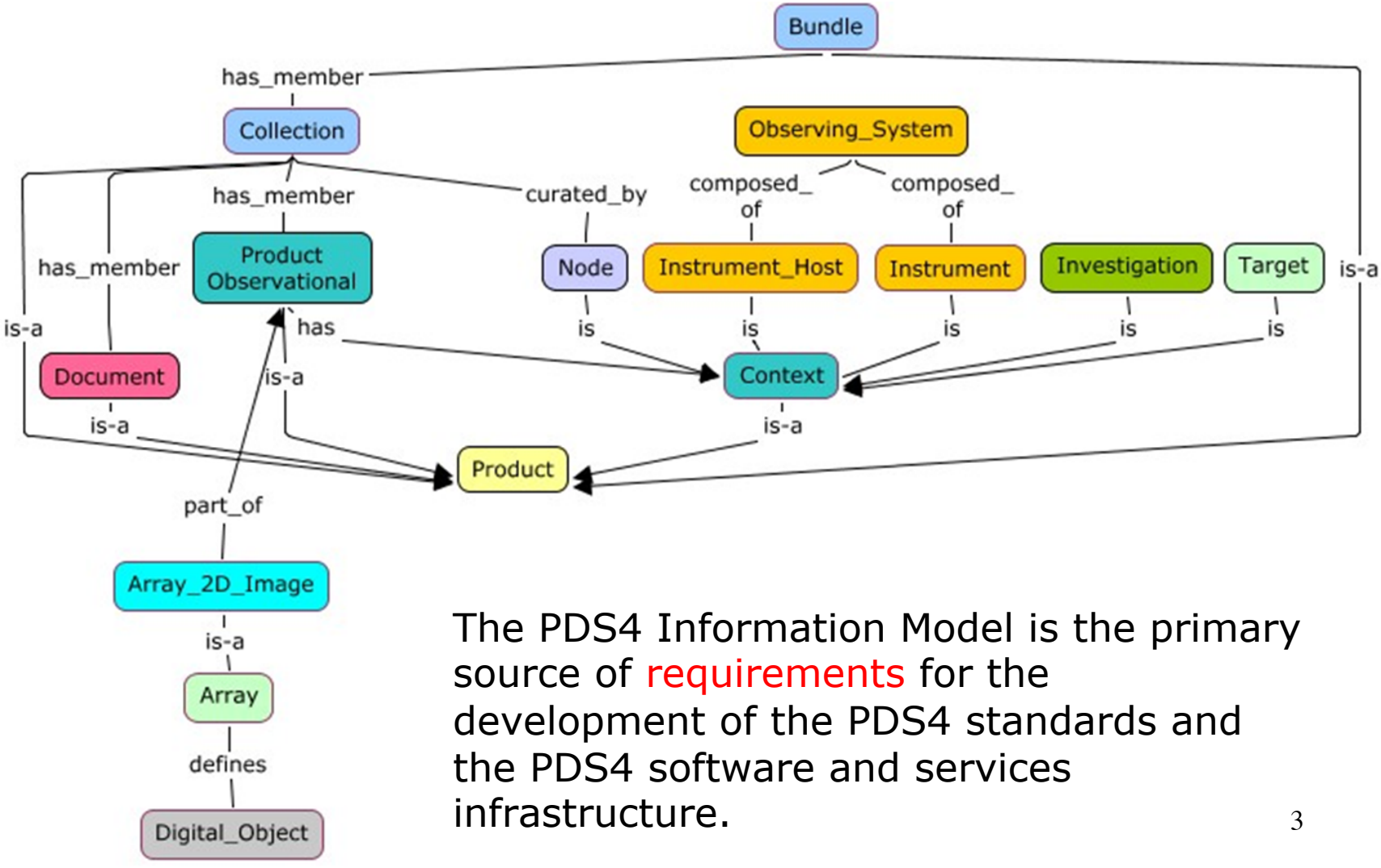
- The IM defines a knowledge base for the Planetary Science community.
- The IM remains independent of the implementation
- Repository obtains the following categories of information
  - **Identification** – allows information object to be discovered and accessed.
  - **Representation/Format** - allows a data object to be interpreted.
  - **Fixity** - ensures the information object has not been unintentionally altered.
  - **Provenance** – essential for authenticity
  - **Context** - describes the environment in which the data object was created.
  - **Reference** - allows the information objects to be referenced

[1] Open Archival Information System (OAIS) Reference Model (ISO 14721)
[2] Metadata Registry Specification (ISO/IEC 11179)
[3] W3C XML (Extensible Markup Language) specifications

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

General Concept Map



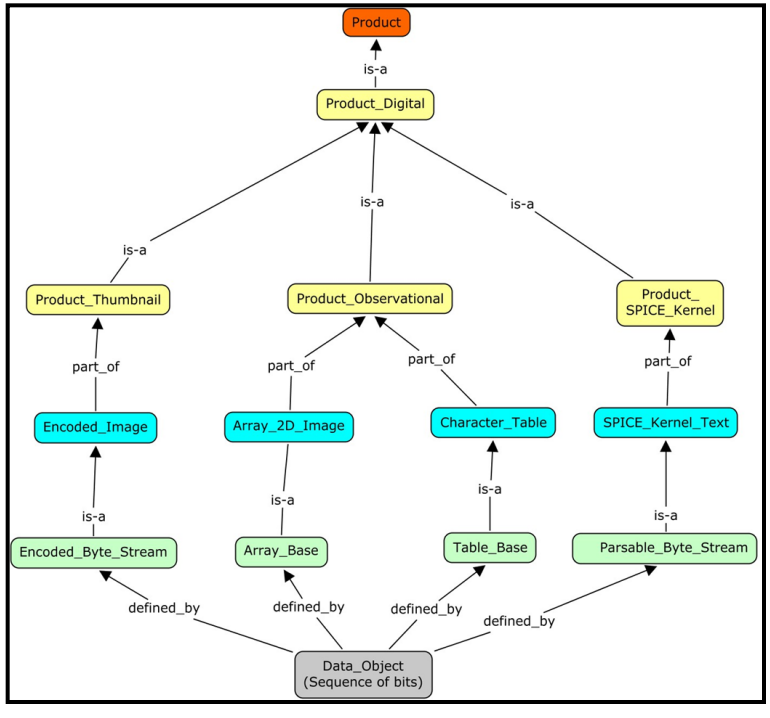
The PDS4 Information Model is the primary source of requirements for the development of the PDS4 standards and the PDS4 software and services infrastructure.

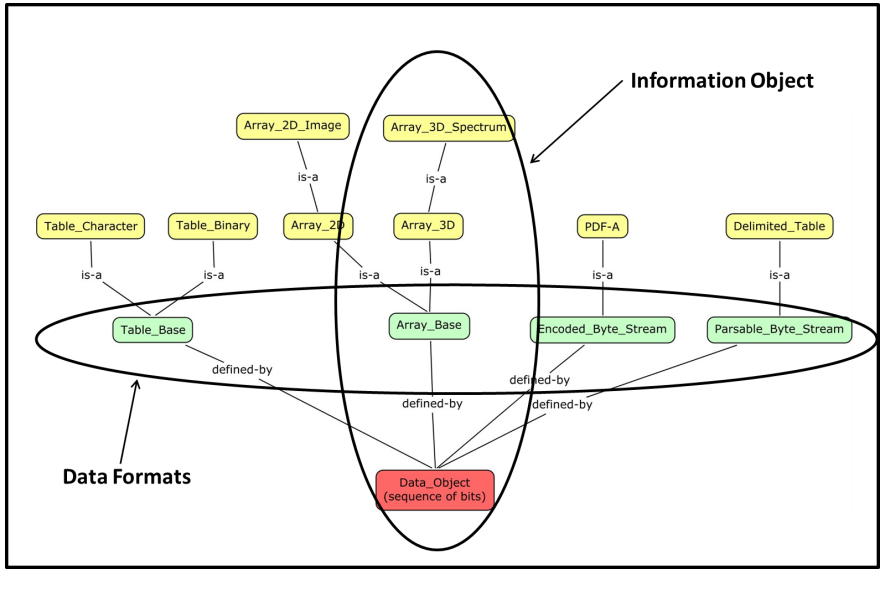
National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Product Concept Map

- Defines and provides templates for capturing
  - data structure (format)
  - context within which the data was captured, processed, and archived
  - the relationships between the data



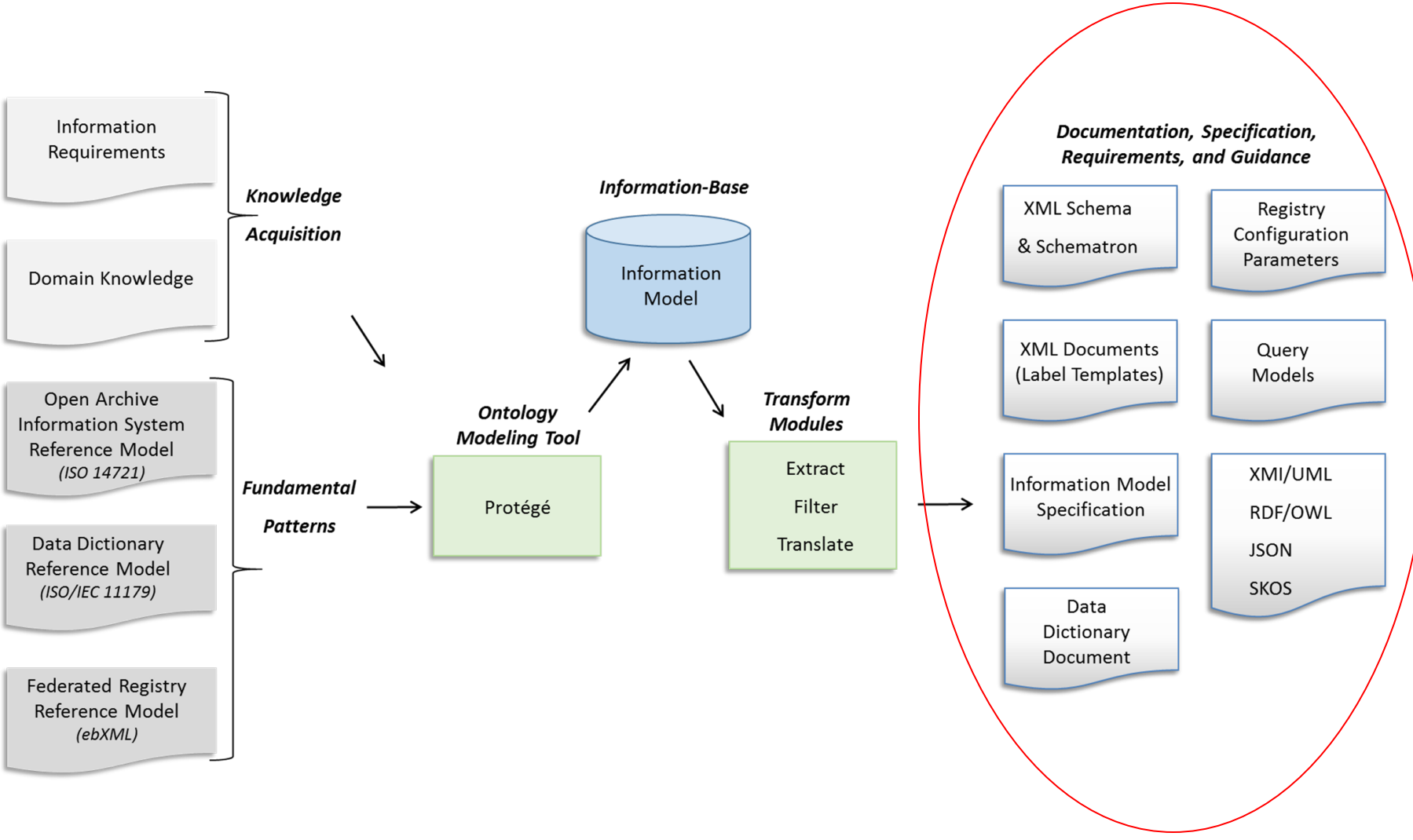


- The single authoritative source for:
  - *The System's Data Standards.*
  - *The System's Data Requirements*
    - registration, search, retrieval, process, analysis, and long-term preservation

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

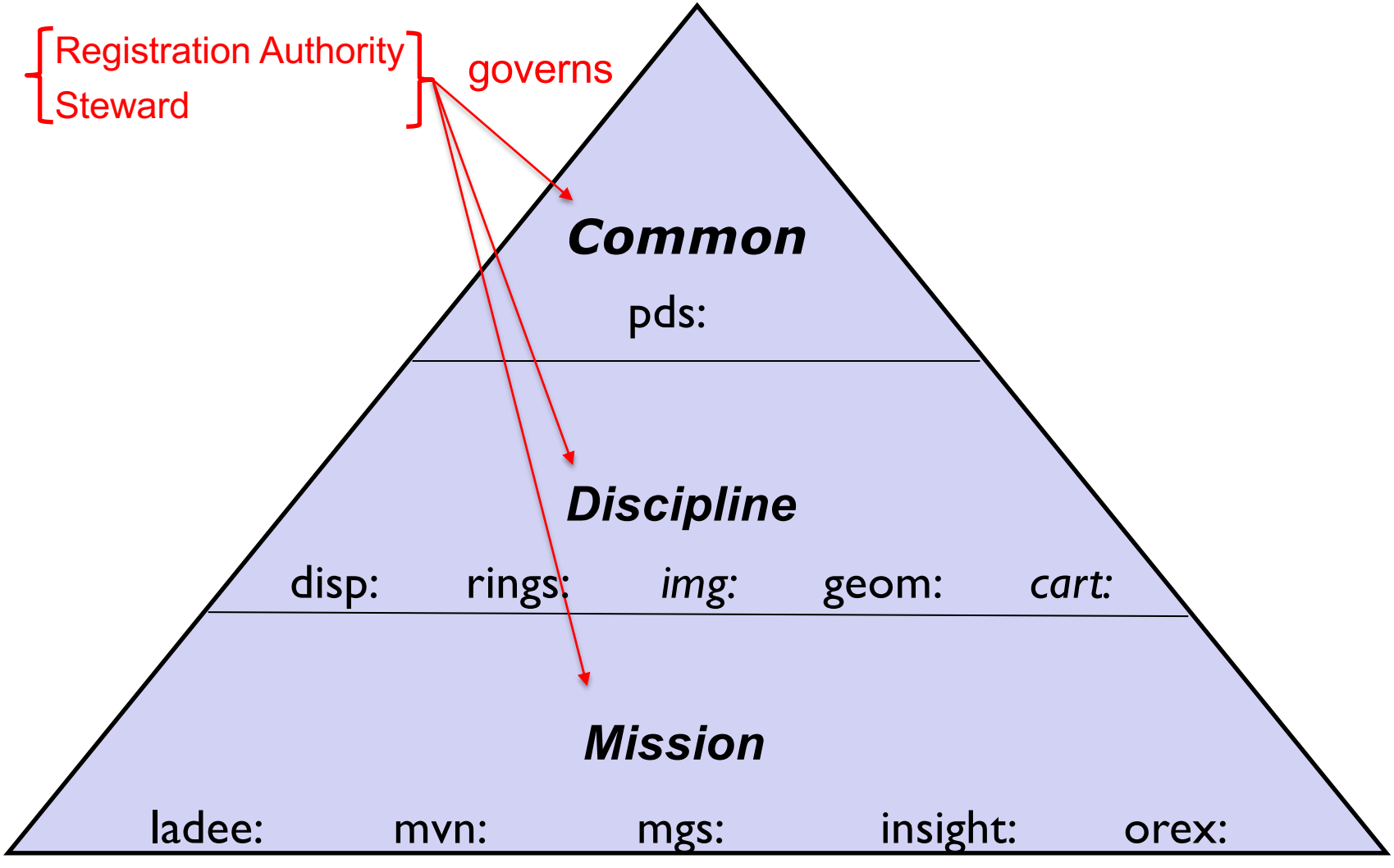
Configuring Software and Services



National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Multi-level Governance



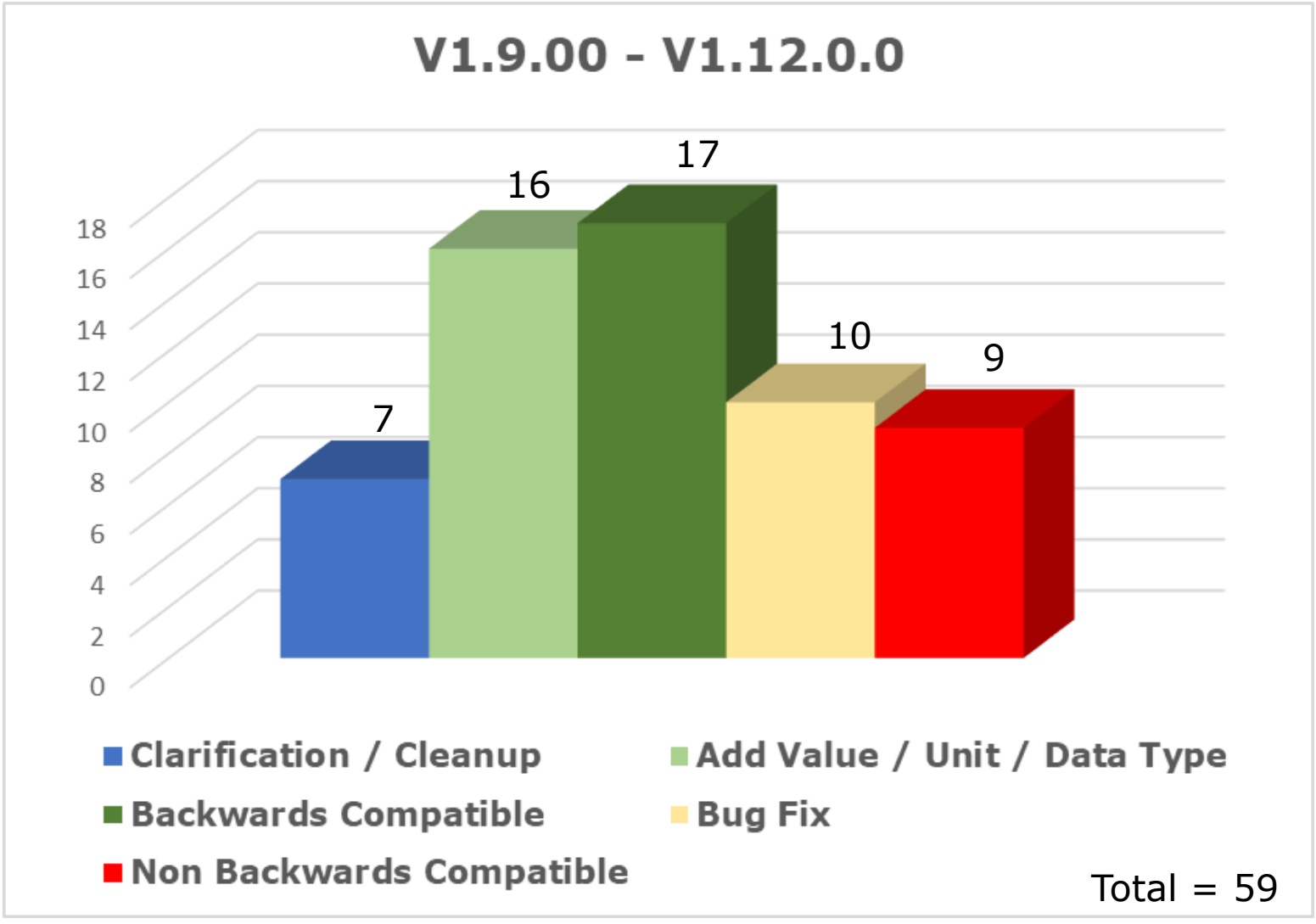
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Controlled Extensions

Common		
	PDS4_PDS_1B10	PDS4 Common
Discipline		
	PDS4_ALT_1000	PPI Node's Alternate
	PDS4_CART_1B10_1930	Imaging Node's Cartography
	PDS4_DISP_1B10	Imaging Node's Display
	PDS4_GEOM_1B10_1700	Geometry
	PDS4_IMG_1B00_1600	Imaging Node
	PDS4_IMG_SURFACE_1B10_1100	Surface Imaging
	PDS4_MSN_SURFACE_1B00_1100	Surface Mission Information
	PDS4_MSN_1B00_1100	Generic Mission
	PDS4_MULTI_1900_1000	multidimensional data
	PDS4_PARTICLE_1900_1100	PPI Node's Particle
	PDS4_PROC_1900	Processing History
	PDS4_RINGS_1800_1500	Rings Node
	PDS4_SP_1C00_1100	Spectral
	PDS4_SPECLIB_1000	Spectral Library
	PDS4_WAVE_1000	PPI Node's Wave
Mission		
	BOPPS_1100	BOPPS
	PDS4_CASSINI_1B00_1200	Cassini
	PDS4_DAWN_1B00_1000	Dawn
	PDS4_INSIGHT_1B00_1850	Insight
	PDS4_JUNO_1900	JUNO
	LADEE_1100	LADEE
	PDS4_MESS_1B00_1020	Messenger
	MGS_1700	*Mars Global Surveyor*
	MPF_1700	*Mars Pathfinder*
	PDS4_MVN_1021	MAVEN
	OSIRIS-Rex _1700	*OSIRIS-Rex*
	PDS4_VG1_1900_1000	VOYAGER 1 (vg1)
	PDS4_VG2_1900_1000	VOYAGER 2 (vg2)

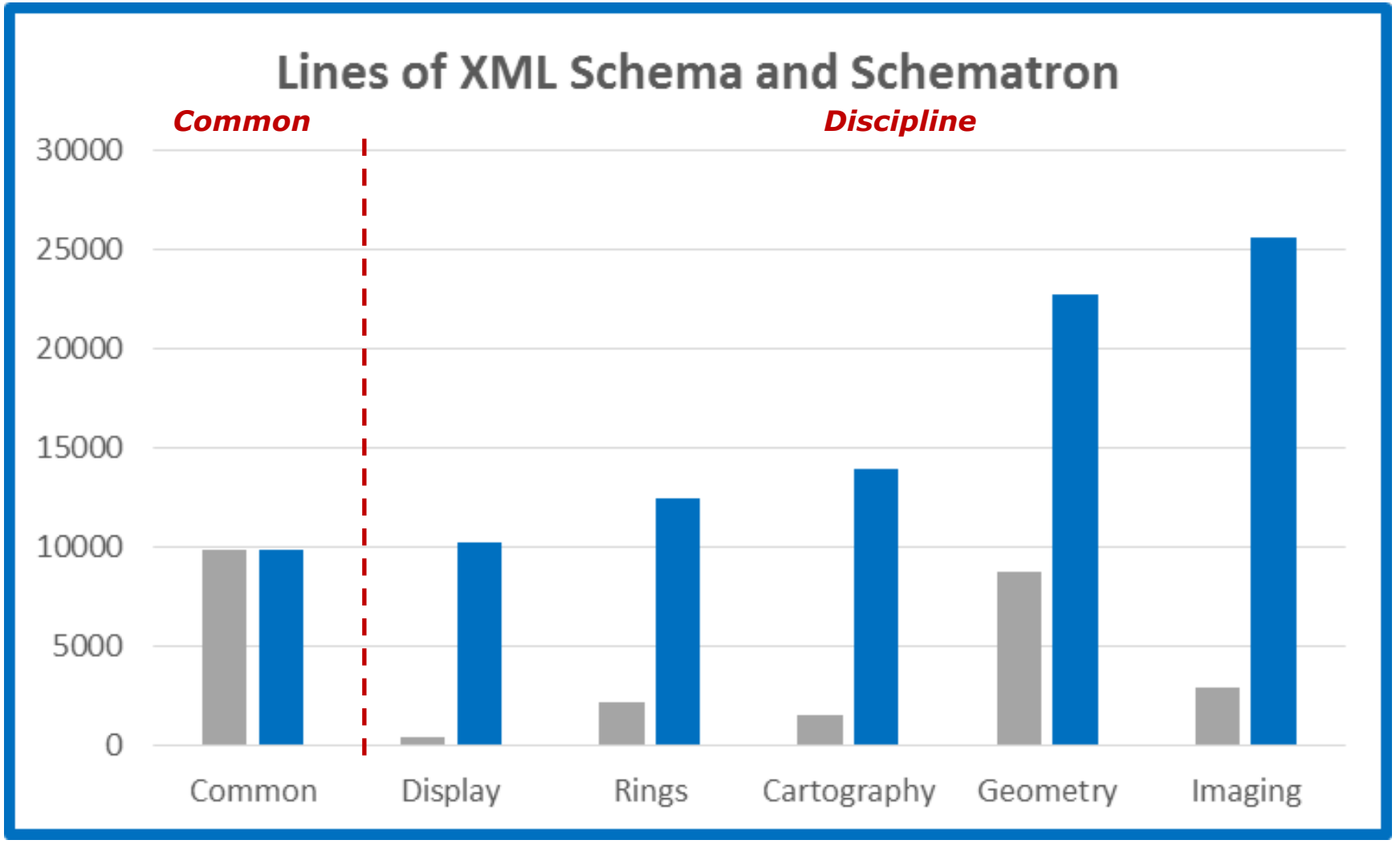
Metrics
Changes to the Common Dictionary



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Auto-generated Validation Code

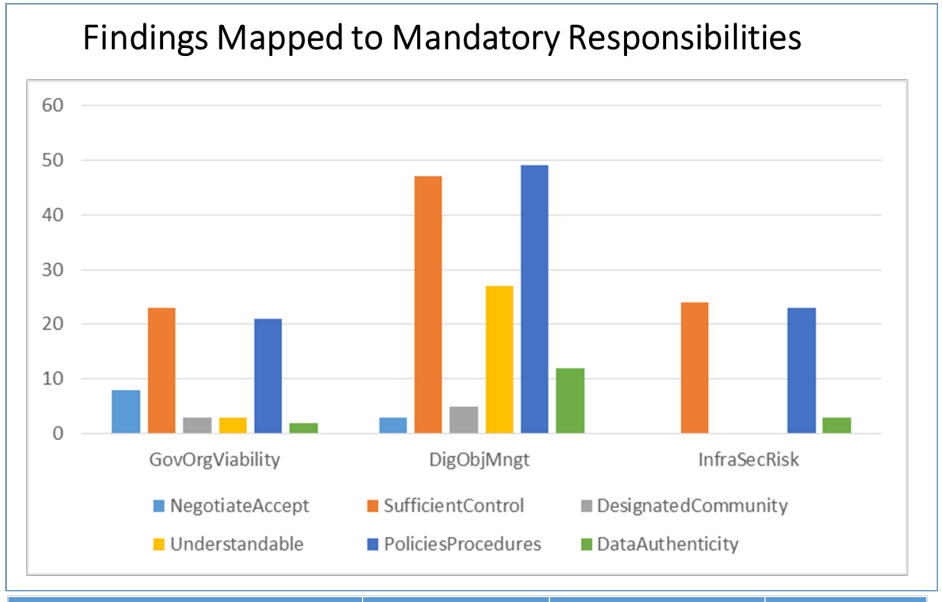


National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Desk Audit – ISO 16363

- The assessment of PDS4 found that 92% of the metrics of the ISO 16363 standard were satisfied.
  - 80% of the metrics for Governance and Organizational Viability
  - 95% of the metrics for Digital Object Management
  - 96% of the metrics for Infrastructure and Security Risk Management.



ISO 16363 Section	Satisfied Metrics	Unsatisfied Metrics	Total Metrics
Governance and Organizational Viability	20	5	25
Digital Object Management	57	3	60
Infrastructure and Security Risk Management	23	1	24
Total	100	9	109

Semantic Technologies

- Under PDS4 all registry objects are first class products.
  - *All products have a Persistent Identifier (PID)*
  - *Named relationships (semantic) are used to relate products using PDS4 PIDs*
    - data, documents, people, software, and contextual objects
  - *Supports Linked Open Data*

- *Evolutionary development*
  - *Adding value by leveraging semantic content*

Conclusions

- The PDS4 Information Model-Driven architecture is meeting the requirements for a "trusted" digital repository for long-term preservation.
  - Maintains the usefulness, reusability, and interoperability of the data.
  - Stays relevant through involvement of domain experts
  - Follows "agile" develop cycle with periodic delivery and testing
- Unleashing the full power of PDS4
  - See "Multimission Labels in PDS", Deen, B., et. al.

Conclusions

- The PDS4 Information Model will scale to meet the increased diversification in science data types and science data processing and analytics
  - Expanding the community through the IPDA.
  - Extensions for:
    - dynamic metadata (updates and transients)
    - targeted relevant searches
    - data analytics
    - semantic reasoning
    - reproducibility

Thank You

Backup

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Why be certified?

- To provide proof that the repository is doing a good job
- Help users make a choice between competitive repositories
- Instils confidence in depositors – so they are more likely to deposit
- Instils confidence in consumers – so they are likely to increase their usage and have greater trust in the information provided
- Instils confidence from upper management
- Instils confidence from funders – i.e. that they are not wasting their money
- Certification under ISO 16363 demonstrates and reinforces the repository's commitment to its mission.

PDS4 Overview

- PDS4 is an information model-driven system architecture
  - *Supports the capture, management, distribution and integration of planetary science data captured in distributed data archives world-wide.*

- The PDS4 Information Model (IM) is the core element of the architecture
  - *Developed using lessons learned from 20 years of archiving Planetary Science Data*
  - *Used accepted standards for information model development*
    - Open Archival Information System (OAIS) Reference Model (ISO 14721)
    - Metadata Registry Specification (ISO/IEC 11179)
    - W3C XML (Extensible Markup Language) specifications.